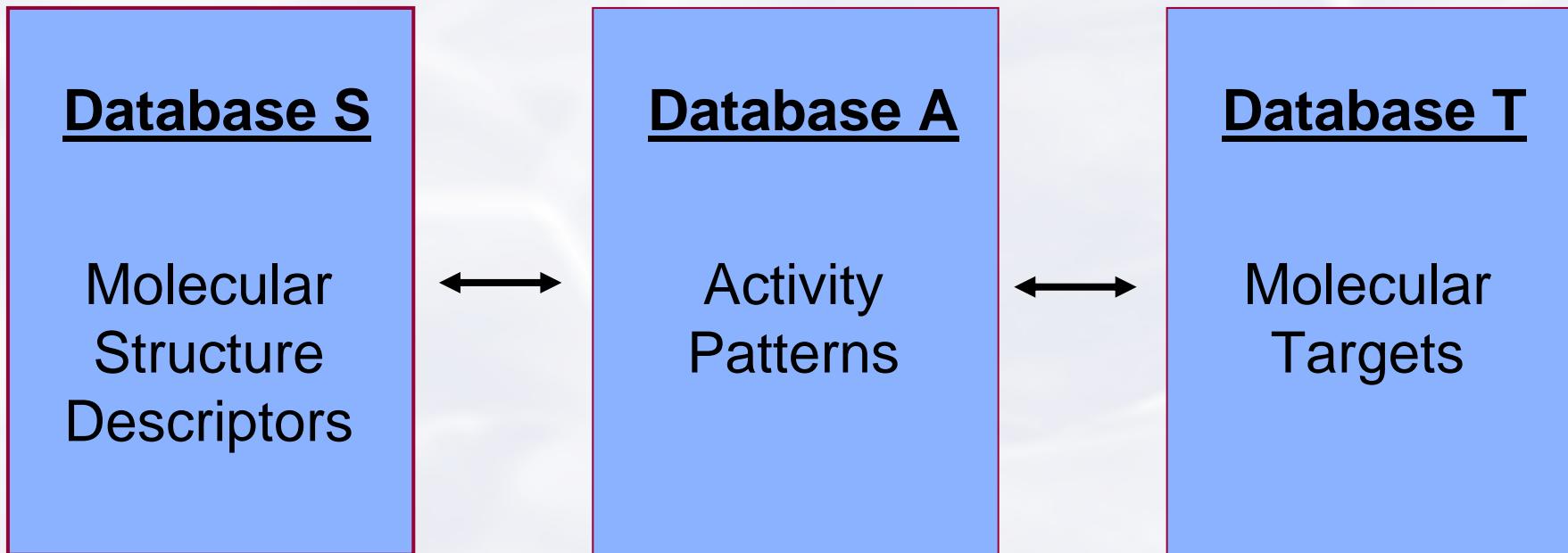


# Understanding toxicity from biological and chemical fingerprints

Chihae Yang  
May 22, 2007

# “Information intensive approach”



- Weinstein, JN, et.al. *Science*, vol 275, January 1997.
- *Pharmacogenomics Journal*, vol 2, 259-271, 2002

# Current reality in toxicity...

Compounds

bioassay data

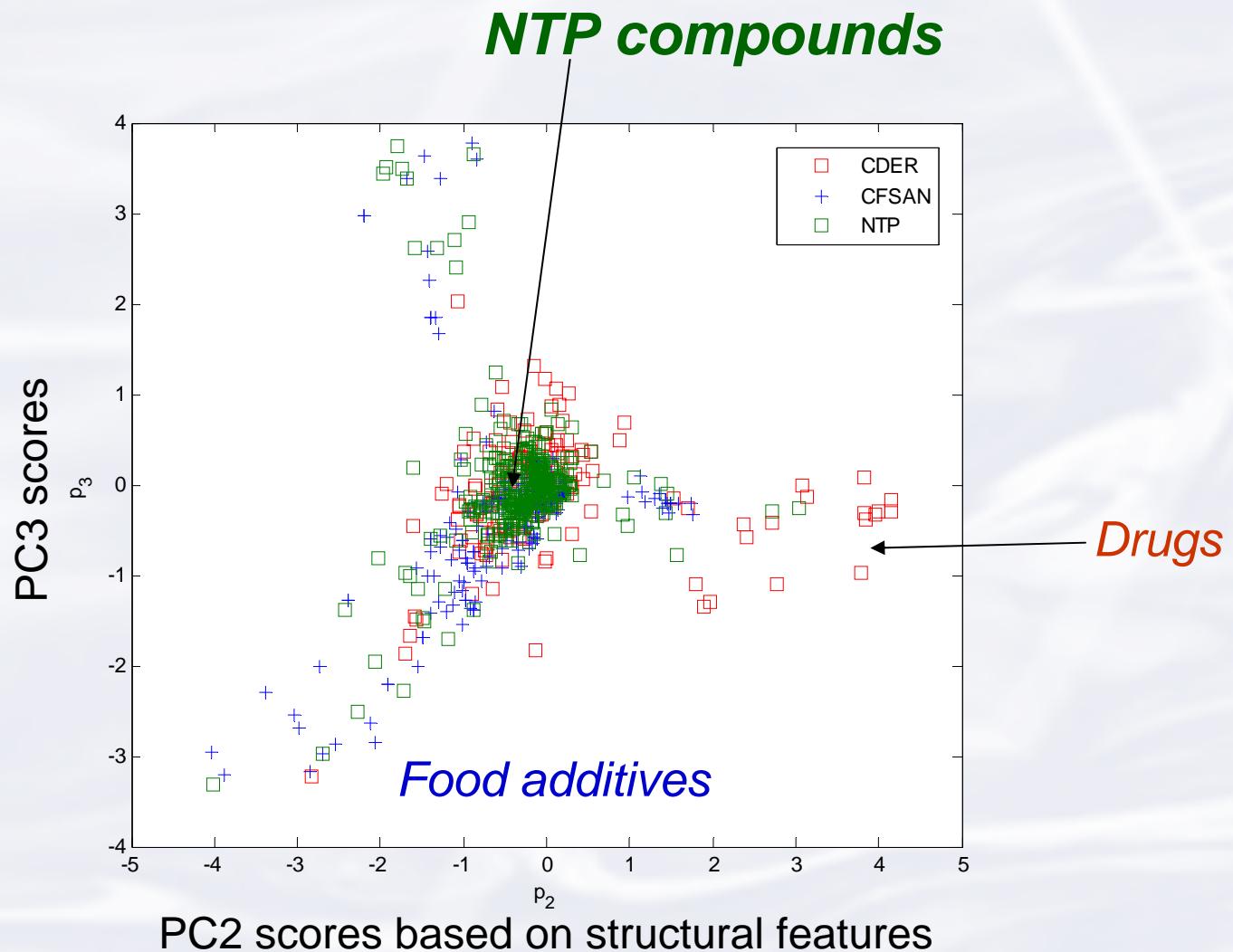
	St	L	S	C	D	D	SH	3T3	H4I	BJ	I	Hek	Hep	Jurk	N2A	SKN	Jurk	MRI	Hek	BJ	I	Hep	Mes	N2A	HU	FDA	NTP	FDA	CPD	CCR	NTP	Sam	RTEC
N'L<15 # #	5	14	1	3	4	1	11	5	7	8	8	14	10	12	9	12	11	19											RTEC				
N'L<15 # #	6	12	2	3	6	2	8	4	7	10	10	14	8	8	12	20	13	16											RTEC				
N'D<15 # #	8	14	1	4	4	3	20	4	7	16	20	12	12	24	19	25	19	26											RTEC				
N'L<15 # #	5	12	2	3	2	2	12	8	4	14	10	10	9	16	10	9	9	27											RTEC				
NTP 15 # #	7	11	3	3	12	1	12	##	8	13	17	14	12	18	11	15	17	15											RTEC				
N'L<15 # #	6	15	2	5	3	1	9	6	8	22	57	9	18	22	17	12	35	13											RTEC				
N'L<15 # #	4	11	2	4	3	1	7	3	6	28	12	10	10	8	11	18	24	20											RTEC				
N'L<15 # #	5	11	2	2	3	1	5	6	5	12	10	16	10	15	9	9	12	17											RTEC				
NTP 16 # #	5	14	13	6	7	2	12	9	9	38	6	17	86	33	58	##	##	37											RTEC				
N'L<16 # #	7	13	2	5	3	1	9	3	11	16	8	12	14	24	18	23	15	27											RTEC				
NTP 16 # #	4	11	2	2	5	2	8	5	8	8	11	13	6	14	10	19	##	29											RTEC				
N'L<16 # #	6	10	2	3	3	1	5	3	6	14	9	12	11	8	16	14	16	26											RTEC				
NTP 16 # #	3	7	2	3	4	2	6	4	4	18	7	9	9	20	7	11	8	13											RTEC				
N'L<16 # #	7	11	3	4	5	2	9	4	9	13	11	11	14	20	13	28	34	35											RTEC				
N'L<16 # #	6	12	2	5	3	2	7	5	7	15	7	10	10	13	19	17	13	29											RTEC				
N'L<16 # #	5	12	2	2	4	2	7	6	8	11	8	8	12	15	6	10	14	19											RTEC				
NTP 16 # #	7	10	3	3	4	1	9	4	8	12	8	8	11	9	8	7	16	19											RTEC				
N'L<16 # #	6	10	2	3	4	1	6	4	3	10	10	8	11	14	7	6	12	19											RTEC				
N'L<16 # #	26	14	5	5	4	1	8	7	13	20	31	10	11	31	10	19	##	21										RTEC					
N'L<16 # #	5	25	2	3	2	1	11	3	5	15	12	13	27	15	10	10	11	24	FDA	FDA	CARC								RTEC				
N'L<17 # #	3	10	2	6	3	1	10	5	6	20	6	10	7	19	12	10	17	23											RTEC				
N'L<17 # #	4	19	30	3	4	2	8	3	9	16	14	11	10	17	22	11	30	23											RTEC				
N'L<17 # #	4	17	3	3	5	2	10	3	8	6	15	11	10	15	15	6	20	25											RTEC				
NTP 17 # #	7	10	4	5	5	3	11	5	6	18	13	11	15	35	20	20	15	36											RTEC				
N'L<17 # #	3	12	2	3	2	1	8	4	9	25	7	11	6	10	9	9	17	18											RTEC				
N'L<17 # #	7	13	1	2	7	1	10	5	6	10	9	12	10	15	10	15	13	17											RTEC				
N'L<17 # #	4	10	2	3	3	2	8	5	4	13	9	10	8	10	9	14	12	29											RTEC				
N'L<17 # #	5	13	2	3	6	2	8	7	4	13	12	15	6	21	13	14	9	21											RTEC				
N'L<17 # #	5	14	2	2	5	2	8	3	4	11	10	6	11	16	13	14	11	36											RTEC				
N'L<17 # #	6	14	2	4	6	2	6	5	7	10	9	8	10	11	12	13	14	28											RTEC				
N'L<17 # #	4	10	2	3	3	2	7	4	6	18	6	16	9	8	14	19	15	26											RTEC				
N'L<18 # #	3	10	2	4	3	2	9	2	6	9	12	7	12	14	6	18	30	19											RTEC				
N'L<18 # #	7	12	2	4	4	1	6	4	5	16	12	11	11	12	12	12	21	20											RTEC				
N'L<18 # #	8	11	2	5	3	2	9	5	9	30	20	15	11	34	30	21	27	39											RTEC				
NTP 18 # #	4	9	2	3	5	2	13	5	5	13	6	10	14	12	13	11	22	20											RTEC				
NTP 18 # #	5	10	2	4	4	1	8	3	6	12	11	7	15	8	13	9	14	14											RTEC				
N'L<19 # #	4	6	2	5	5	4	6	4	4	7	6	11	11	9	8	18	10	11											RTEC				
N'L<19 # #	6	11	2	2	4	2	6	6	6	19	8	10	12	15	10	13	12	19											RTEC				

toxicity data

# NTP high throughput screening

- A method to prioritize the selection of chemicals for NTP bioassays
  - Test chemical nomination procedure
    - rank or rule out “most likely toxic” and “most likely not toxic” compounds
- An approach to potentially obtain insights for mode of actions
- Predictive toxicology?

# NTP HTS 1408 Chemicals



# NTP HTS Biological assays

- Activity: NCGC 10 strains
  - Caspases - 3T3, BJ, SHSY5, H4Ile ,Hek293, HepG2, HUVEC, Jurkat, N2A
  - I kB signaling, JNK Alpha
- Viability: Fred (7 strains), NCGC (13 strains)
  - 3T3, BJ, SHSY5, H4Ile ,Hek293, HepG2, HUVEC, Jurkat, N2A
- “Other assays” – NCGC (4 assays)
  - SKNSH, MRC5, Renal, Mesenchymal

# Toxicity data used for profiling NTP HTS compounds

- Carcinogenicity
  - Rodent bioassay data
- Genetic toxicity
  - Salmonella strains
  - Mammalian mutagenesis – mouse lymphoma, HPRT
  - Clastogenicity – in vitro CA, In vivo MN...
- Rodent acute toxicity
- Endocrine disruptors

# Sparse toxicity data space

NLS1#	D	C	I	D	SH53T3	H4i	BJ-r	Hek	Hep	JHU	Jurk	N2a	SKN	Jurk	MR	Hek	BJ-r	Hep	Mes	N2a	HU	FDA	CPD	CCR	NTP	Sam	RTEI	
NLS1#	5	14	1	3	4	1	11	5	7	8	8	14	10	12	9	12	11	19										
NLS1#	6	12	2	3	6	2	8	4	7	10	10	14	8	8	12	20	13	16	NTP	CCRS_GEN	RTEI							
ND1#	8	14	1	4	4	3	20	4	7	16	20	12	12	24	19	25	19	26	NTP	CCP_NTP_Sam	RTEI							
NLS1#	5	12	2	3	2	2	12	8	4	14	10	9	16	9	9	9	27	NTP	CCP_NTP_Sam	RTEI								
NTP1#	7	11	3	3	12	1	12	#	8	13	17	14	12	16	11	15	17	15	NTP	CCP_NTP_Sam	RTEI							
NLS1#	6	15	2	5	3	1	9	6	8	22	57	9	18	22	17	12	35	13	NTP	CCP_NTP_Sam	RTEI							
NLS1#	4	11	2	4	3	1	7	3	6	28	12	10	10	8	11	18	24	20	NTP	CCP_NTP_Sam	RTEI							
NLS1#	5	11	2	2	3	1	5	6	5	12	10	16	10	15	9	9	12	17	NTP	CCP_NTP_Sam	RTEI							
NTP1#	5	14	13	6	7	2	12	9	9	38	6	17	86	33	58	#	#	37	NTP	CCRS_GEN	RTEI							
NLS1#	7	13	2	5	3	1	9	3	11	16	8	12	14	24	18	23	15	27	NTP	CCRS_GEN	RTEI							
NTP1#	4	11	2	2	5	2	8	5	8	8	13	6	14	10	19	10	19	#	29	NTP	CCRS_GEN	RTEI						
NLS1#	6	10	2	3	3	1	5	3	6	14	9	12	11	8	16	14	16	26	NTP	CCRS_GEN	RTEI							
NTP1#	3	7	2	3	4	2	6	4	4	18	7	9	20	7	11	8	13	NTP	CCP_NTP_Sam	RTEI								
NLS1#	7	11	3	4	5	2	9	4	9	13	11	11	14	20	13	28	34	35	NTP	CCP_NTP_Sam	RTEI							
NLS1#	6	12	2	5	3	2	7	5	7	15	7	10	10	13	19	17	13	29	NTP	CCP_NTP_Sam	RTEI							
NLS1#	5	12	2	2	4	2	7	6	8	11	8	8	12	15	6	10	14	19	NTP	CCRS_GEN	RTEI							
NTP1#	7	10	3	3	4	1	9	4	8	12	8	8	11	9	8	7	16	19	NTP	CCRS_GEN	RTEI							
NLS1#	6	10	2	3	4	1	6	4	3	10	10	8	11	14	7	6	12	19	NTP	CCRS_GEN	RTEI							
NLS1#	26	14	5	5	4	1	8	7	13	20	31	10	11	31	10	19	19	#	21	NTP	CCAR	FDA						
NLS1#	5	25	2	3	2	1	11	3	5	15	12	13	27	15	10	10	11	24	NTP	CCRS_GEN	RTEI							
NLS1#	3	10	2	6	3	1	10	5	6	20	6	10	7	19	12	10	17	23	NTP	CCP_NTP_Sam	RTEI							
NLS1#	4	19	30	3	4	2	8	3	9	16	14	11	10	17	22	11	30	23	NTP	CCRS_GEN	RTEI							
NLS1#	4	17	3	3	5	2	10	3	8	6	15	11	10	15	15	6	20	25	NTP	CCP_NTP_Sam	RTEI							
NTP1#	7	10	4	5	5	3	11	5	6	18	12	11	15	35	20	20	15	36	NTP	CCRS_GEN	RTEI							
NLS1#	3	12	2	3	2	1	8	4	9	25	7	11	6	10	9	9	17	18	NTP	CCRS_GEN	RTEI							
NLS1#	7	13	1	2	7	1	10	5	6	10	9	12	10	15	10	15	13	17	NTP	CCRS_GEN	RTEI							
NLS1#	4	10	2	3	3	2	8	5	4	13	9	10	8	10	9	14	12	29	NTP	CCRS_GEN	RTEI							
NLS1#	5	13	2	3	6	2	8	7	4	13	12	15	6	21	13	14	9	21	NTP	CCRS_GEN	RTEI							
NLS1#	5	14	2	2	5	2	8	3	4	11	10	6	11	16	13	14	11	36	NTP	CCP_NTP_Sam	RTEI							
NLS1#	6	14	2	4	6	2	6	5	7	10	9	8	10	11	12	13	14	28	NTP	CCP_NTP_Sam	RTEI							
NLS1#	4	10	2	3	3	2	7	4	6	18	6	16	9	8	14	19	15	26	NTP	CCP_NTP_Sam	RTEI							
NLS1#	3	10	2	4	3	2	9	2	6	9	12	7	12	14	6	18	30	19	NTP	CCRS_GEN	RTEI							
NLS1#	7	12	2	4	4	1	6	4	5	16	12	11	11	12	12	12	21	20	NTP	CCRS_GEN	RTEI							
NLS1#	8	11	2	5	3	2	9	5	9	30	20	15	11	34	30	21	27	39	NTP	CCP_NTP_Sam	RTEI							
NTP1#	4	9	2	3	5	2	13	5	5	13	6	10	14	12	13	11	22	20	NTP	CCP_NTP_Sam	RTEI							
NTP1#	5	10	2	4	4	1	8	3	6	12	11	7	15	8	13	9	14	14	NTP	CCP_NTP_Sam	RTEI							
NLS1#	4	6	2	5	5	4	6	4	4	7	6	11	11	9	8	18	10	11	NTP	CCRS_GEN	RTEI							
NLS1#	6	11	2	2	4	2	6	6	6	19	8	10	12	15	10	13	12	19	NTP	CCP_NTP_Sam	RTEI							

## Toxicity endpoints

### In vivo micronucleus

### In vitro chromosome aberration

### Mammalian mutagenesis

### Salmonella mutagenesis

structure counts

905

318

45

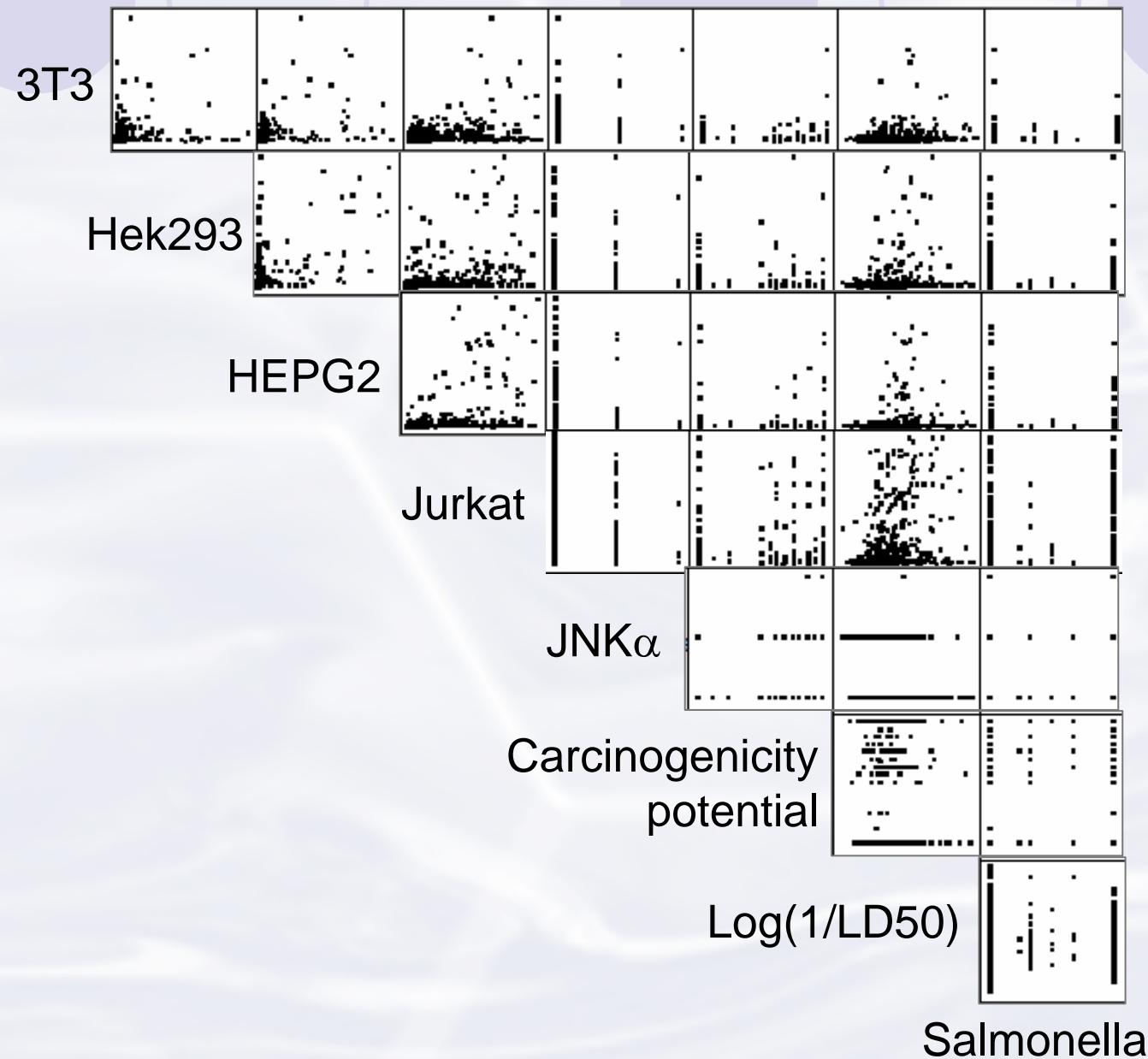
401

314

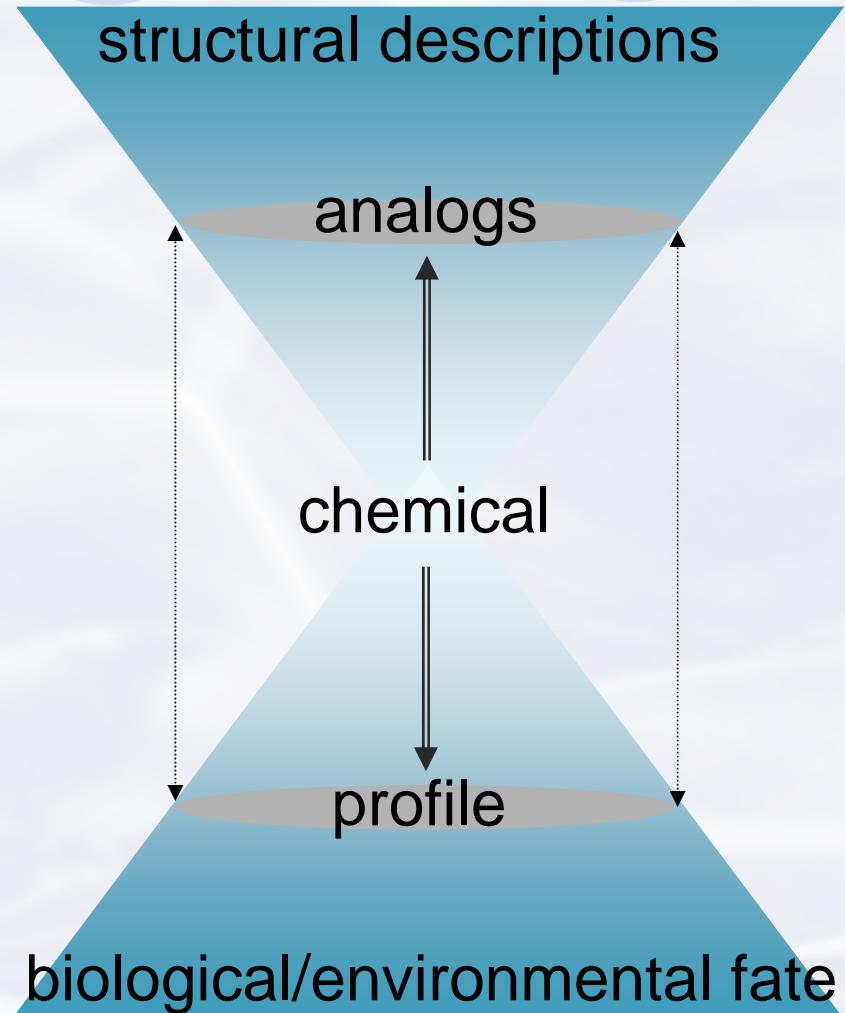
942

total 1340 structures

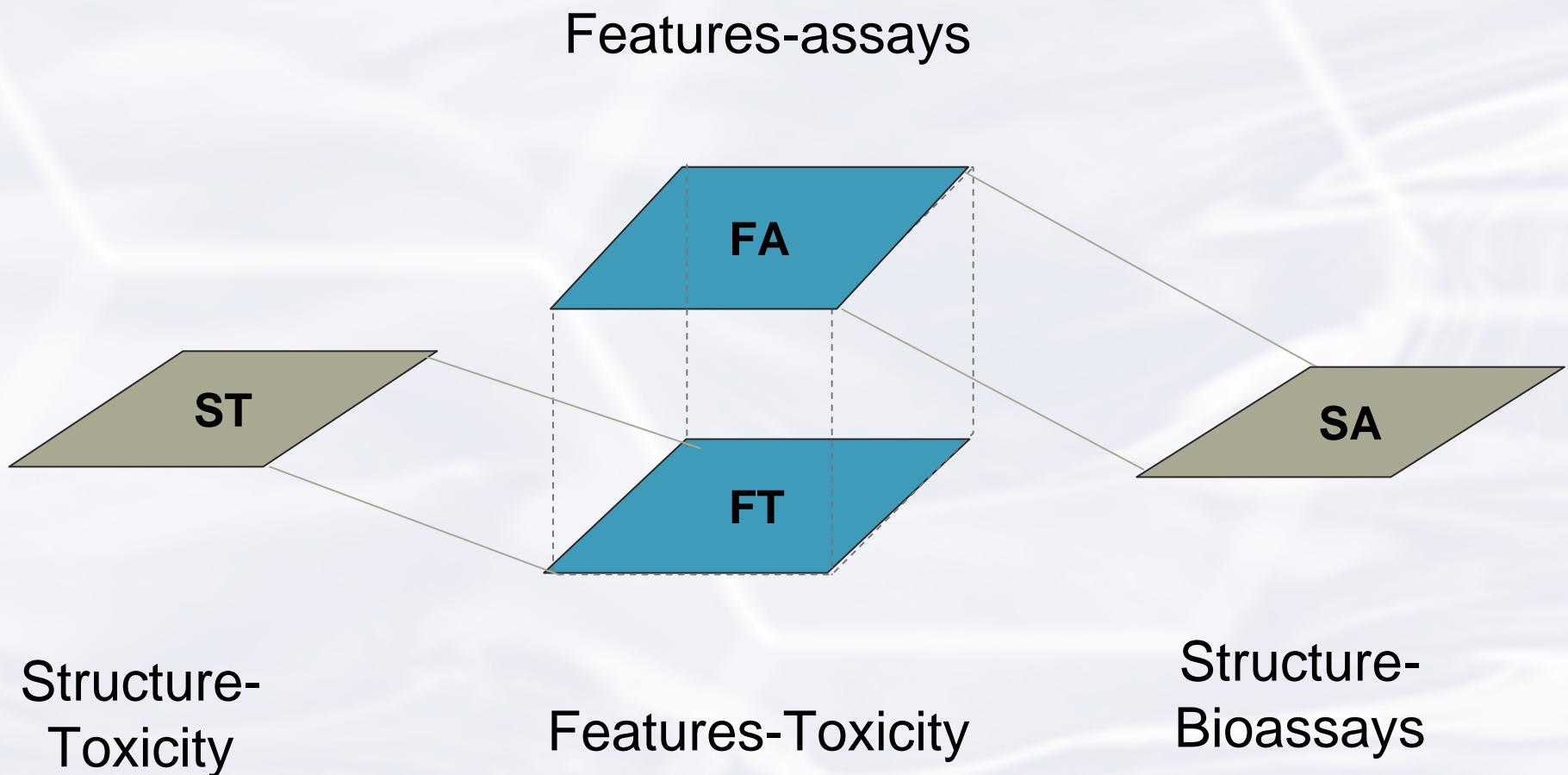
# Compound level correlations



# Linking biology to chemistry



# Features dimension as a link

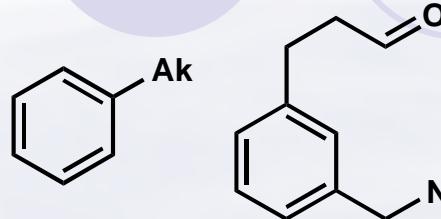


# Features...

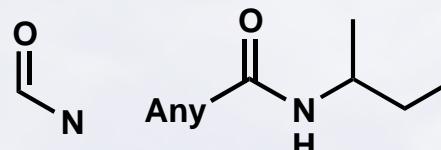
- Chemical fingerprints
  - Structural features
- Calculated descriptors
- Physicochemical properties
- ...

# Representing structures with Leadscope molecular descriptors

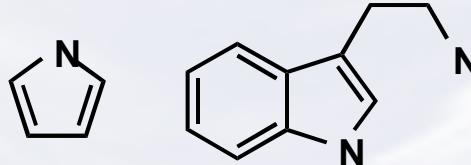
Benzenes



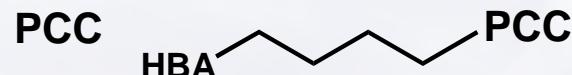
Functional groups



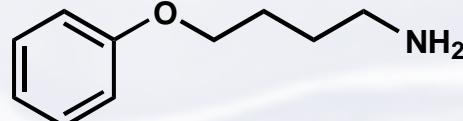
Heterocycles



Pharmacophores



Spacers



User defined features

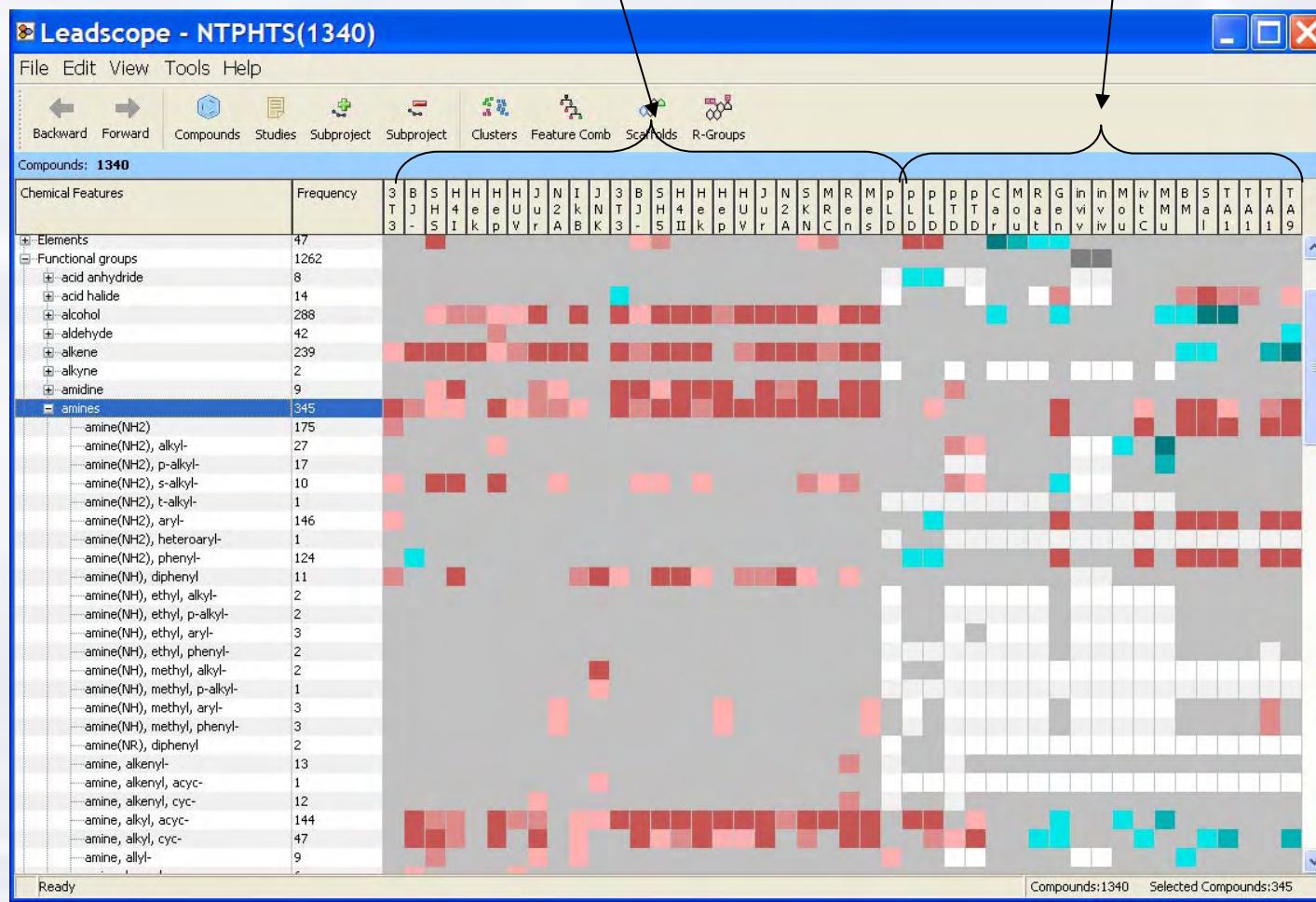
knowledge addition

# Profiling structures with features, bioassays and toxicity endpoints

bioassay data

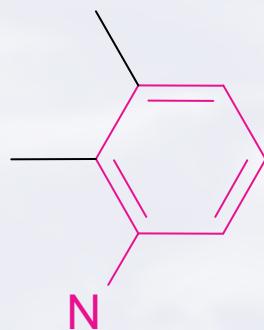
toxicity data

Features

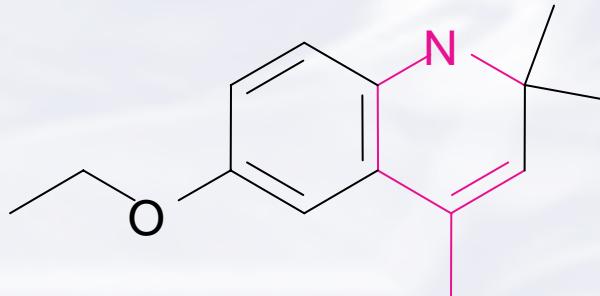


# Feature selection

- 158 features are selected
  - Z-scores are used to distinguish activities
  - Differentiating features per endpoint
  - Highest variances across the endpoints

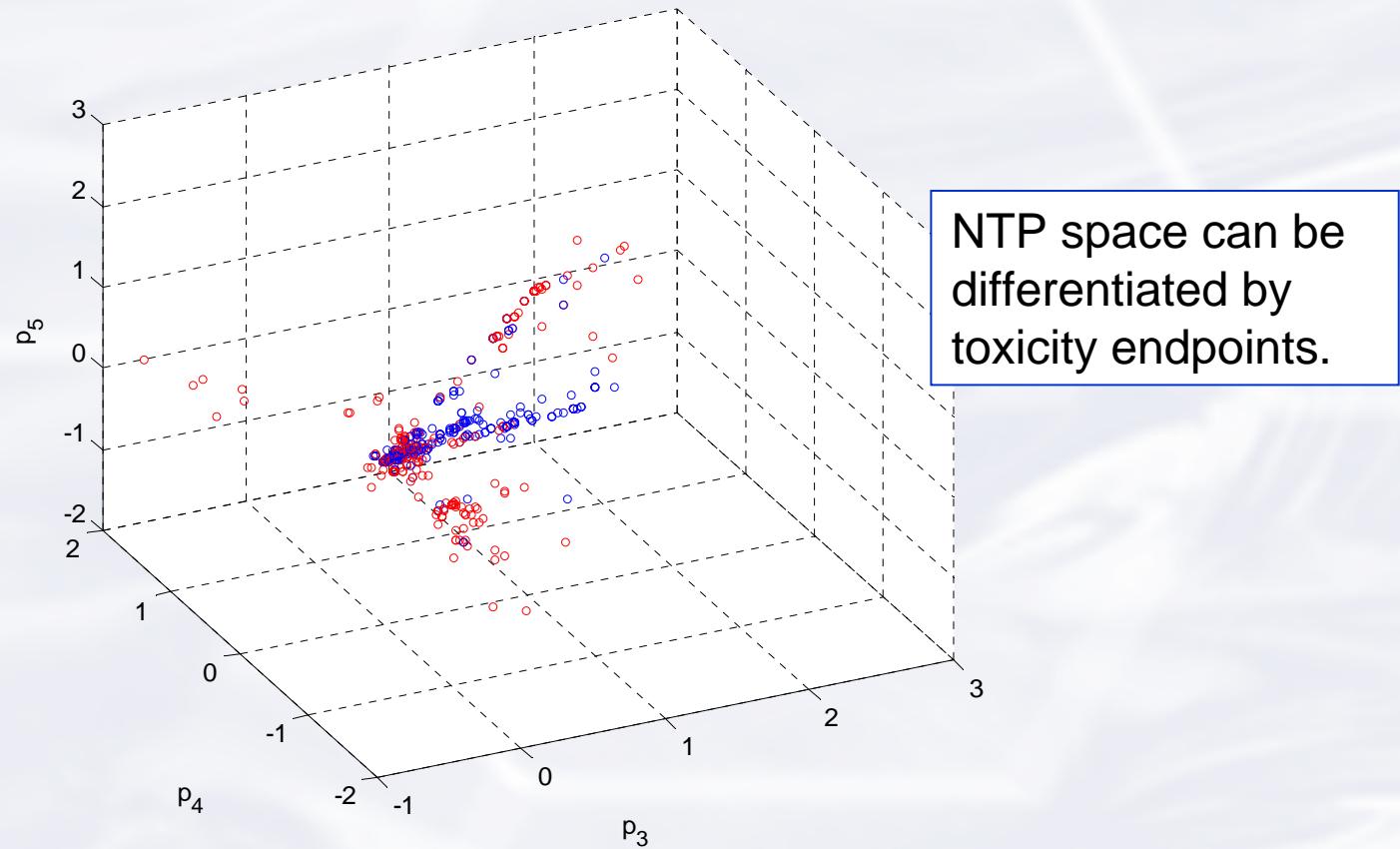


Salmonella:  $Z = 6.9$   
BJ range:  $Z = -2.0$   
Juracat range:  $Z = -1.0$

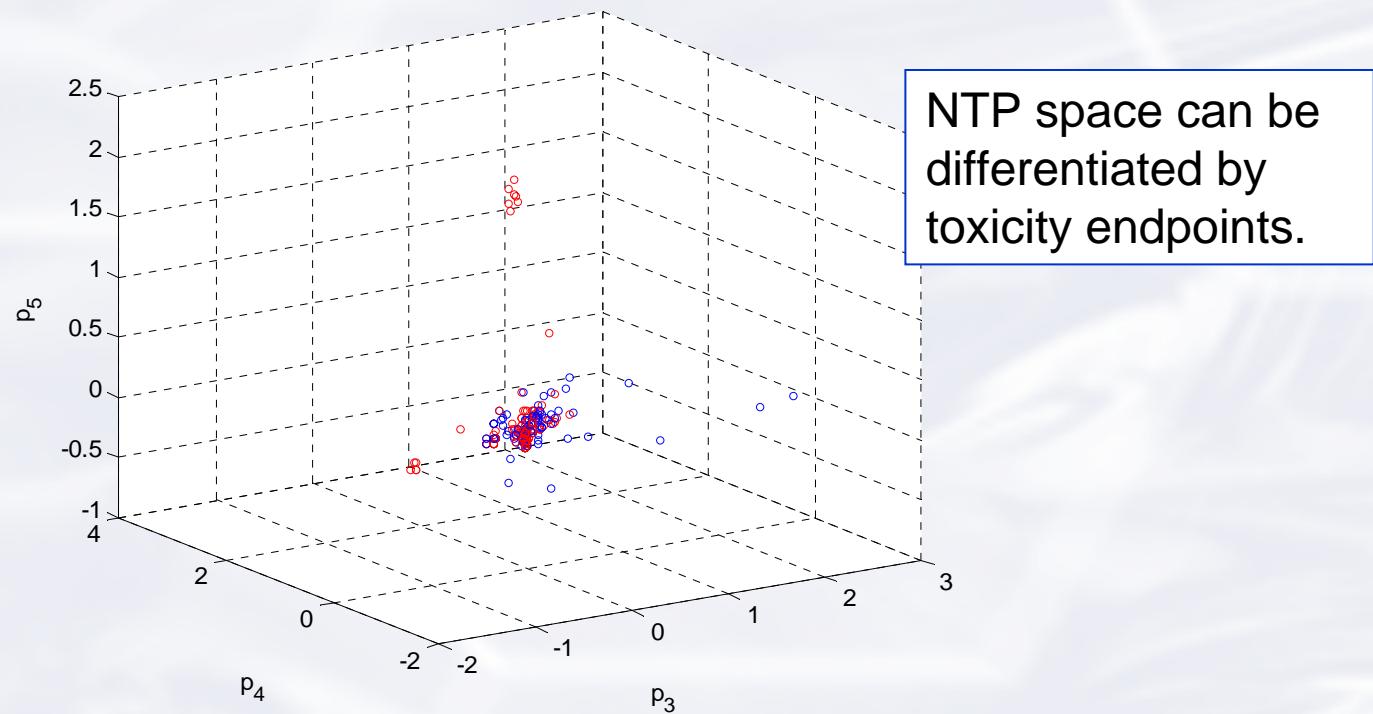


Salmonella:  $Z = -2.1$   
BJ range:  $Z = 4.2$   
Juracat range:  $Z = 4.2$

# Principal Components: discrimination of NTP chemical space by mutagenic potential



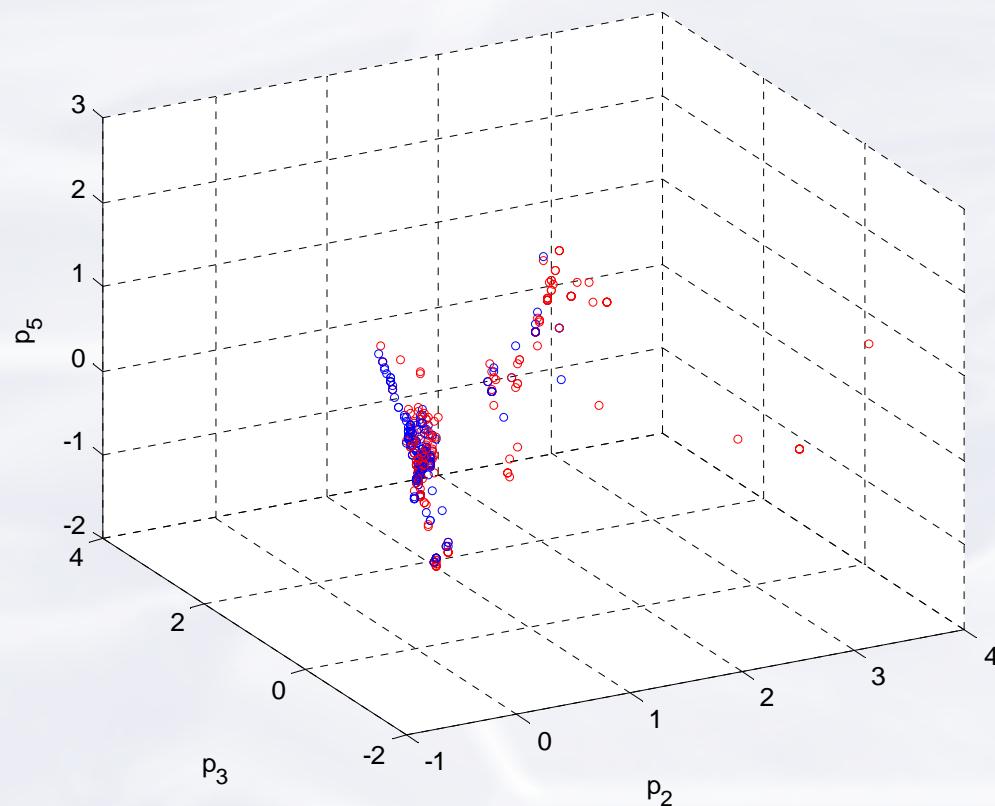
# Principal Components: discrimination of NTP chemical space by carcinogenic potential



Principal component scores:

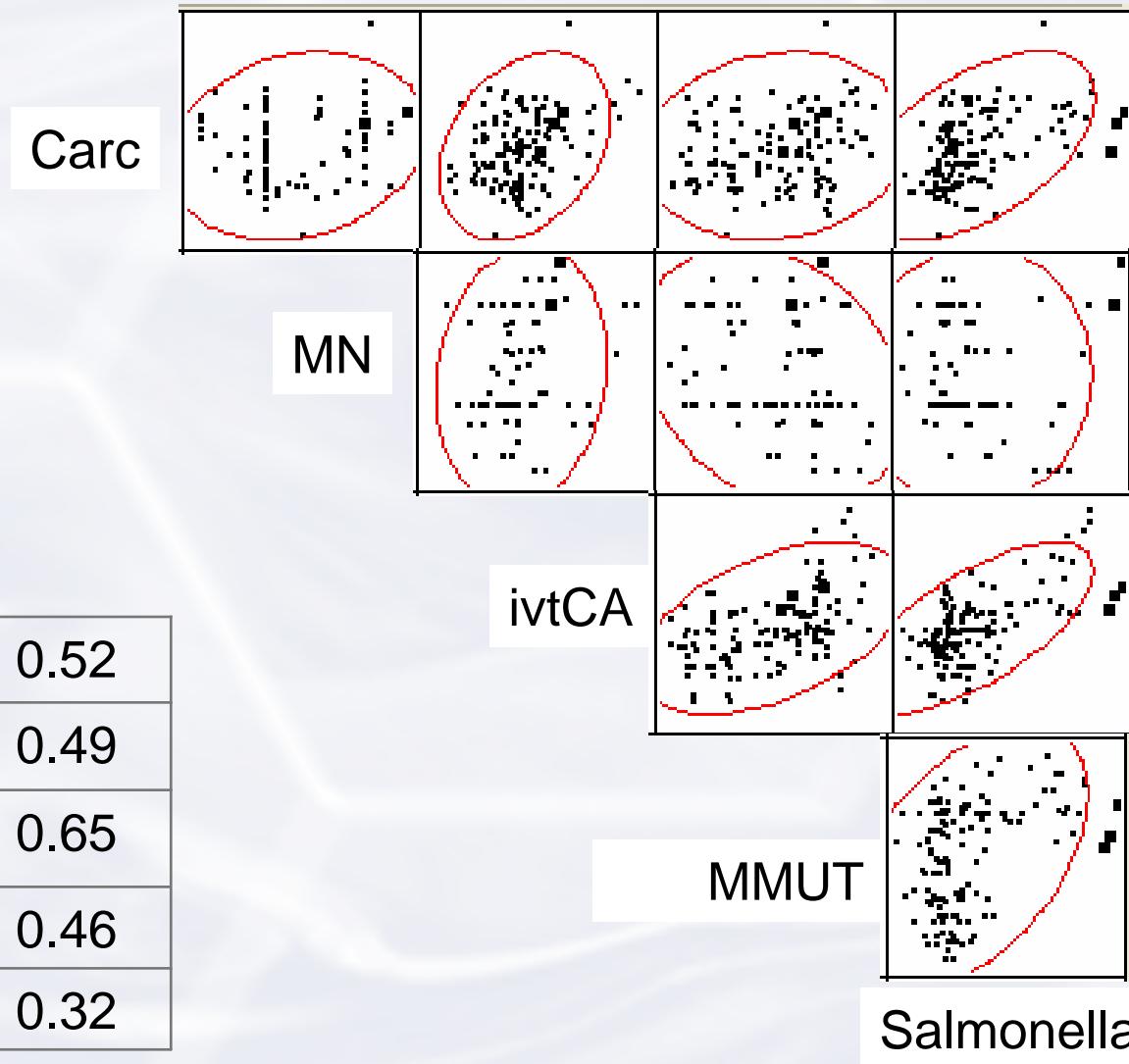
red=Carc positive ( $p>0.5$ ), blue=Carc negative ( $p\leq 0.5$ )

# Principal Components – discrimination of NTP chemical space by acute toxicity

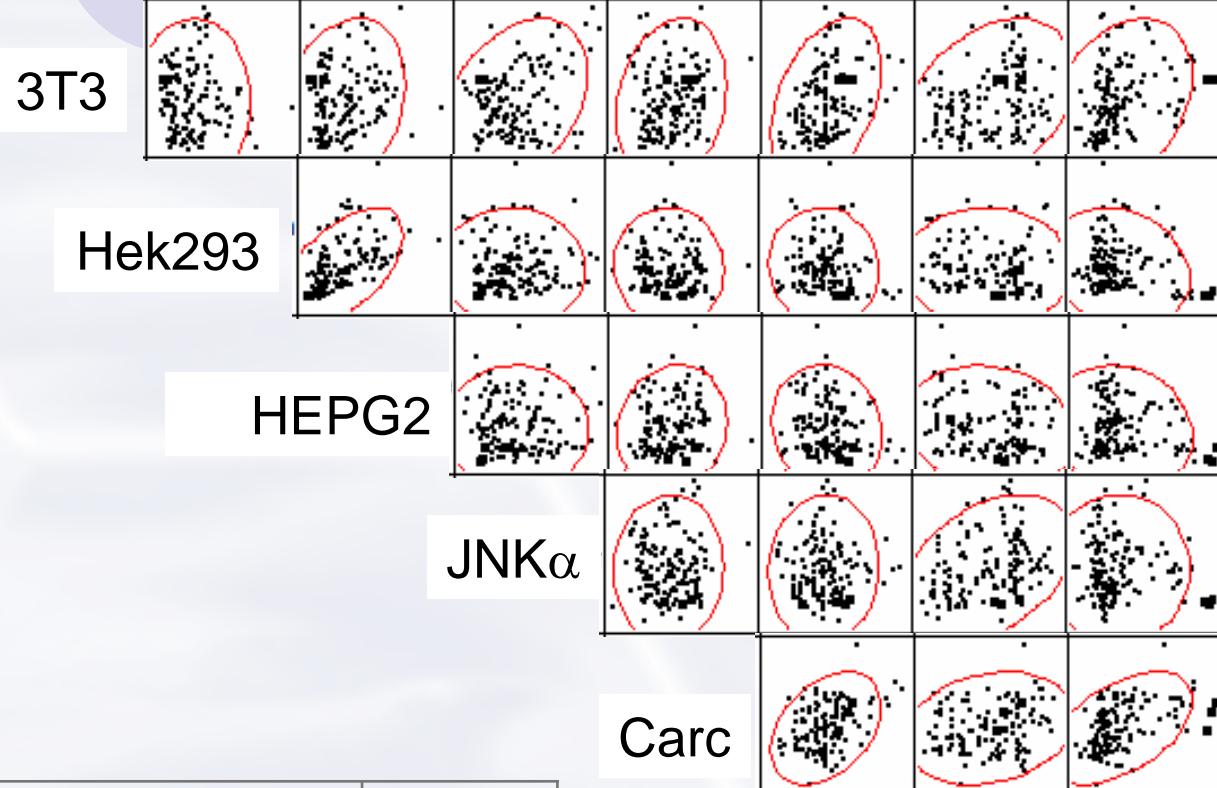


Principal component scores: features  
(red= $p\text{LD}50\text{Rodent} > -0.84$ , blue= $p\text{LD}50\text{Rodent} \leq -0.84$ )

# Correlation between genetic toxicity and carcinogenicity



# 3T3 and JNK with mutagenesis...

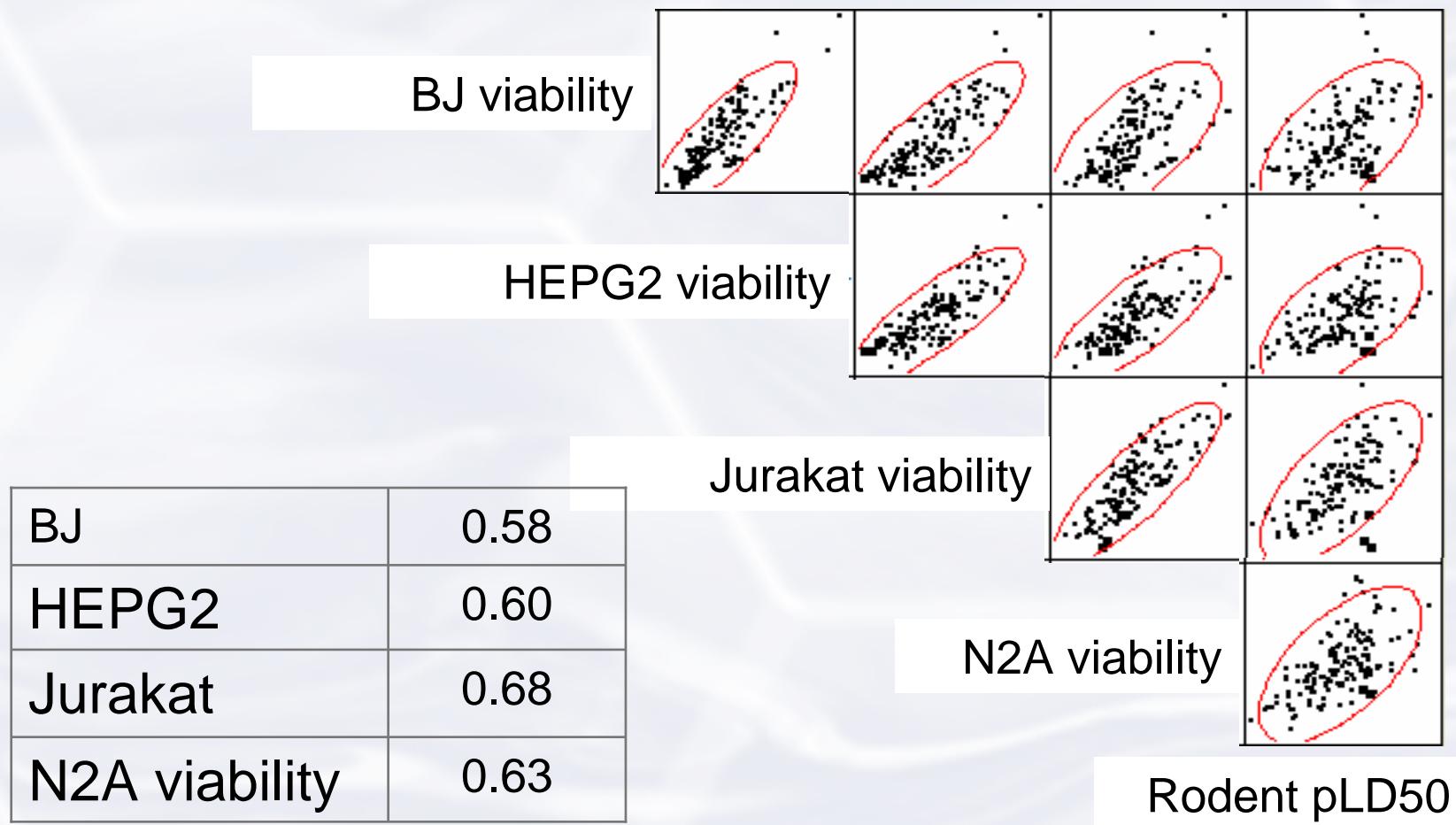


Outlier features:  
• Ar-Nitro  
• Ar-amino(hetero)

Salmonella – 3T3	0.32
ivtCA – 3T3	0.39
MMut – 3T3	0.32
MMut – JNK $\alpha$	0.38

# Viability and acute toxicity

Outlier features: carbonyl halomethyl, imine, iminomethyl



# PLS model for acute toxicity

- Predictors: cell viability of 3T3, BJ, SHSY5, H4Ile, Hek293, HepG2, HUVEC, Jurkat, N2A
- Response variable: pLD50 of rodent for 905 compounds
- Results:
  - at a compound level with 905 observations
    - $Q^2 = 0.122$  (with 2 PLS factors)
  - at a feature level with 158 observations
    - $Q^2 = 0.576$  (with 2 PLS factors)

# Summary

- Correlations need to be at “features” level, not at compound level.
- Chemoinformatics methods may guide compound and bioassay selections.
- Future work
  - Construct a knowledge base of calculated descriptors stratified for mechanisms
  - Develop algorithms for mechanism driven feature selections

# Acknowledgement

- Collaborators
  - Ann Richard, EPA
  - Jennifer Fostel, NIEHS
  - J.F. Rathman, the Ohio State University
  - J. Gasteiger, Molecular Networks
- Leadscope team